Probabilistic Temporal Networks: A unified framework for reasoning with Time and Uncertainty[1]

Eugene Santos Jr.

Department of Computer Science and Engineering

University of Connecticut

Storrs, CT 06269-3155

eugene@cse.uconn.edu

Joel D. Young

Department of Electrical and Computer Engineering

Air Force Institute of Technology

Wright-Patterson AFB, OH 45433-7765

jdyoung@afit.af.mil

[1]This research was supported in part by AFOSR Project #940006.

Abstract

Complex real-world systems consist of collections of interacting processes/events. These processes change over time in response to both internal and external stimuli as well as to the passage of time itself. Many domains such as real-time systems diagnosis, story understanding, and financial forecasting require the capability to model complex systems under a unified framework to deal with both time and uncertainty. Current models for uncertainty and current models for time already provide rich languages to capture uncertainty and temporal information respectively. Unfortunately, these semantics have made it extremely difficult to unify time and uncertainty in a way which cleanly and adequately models the problem domains at hand. Existing approaches suffer from significant trade offs between strong semantics for uncertainty and strong semantics for time. In this paper, we explore a new model, the Probabilistic Temporal Network, for representing temporal and atemporal information while fully embracing probabilistic semantics. The model allows representation of time constrained causality, of when and if events occur, and of the periodic and recurrent nature of processes.

Keywords: temporal reasoning, probabilistic reasoning, abduction, linear constraint satisfaction, cyclicity

1 Introduction

In the evolution of expert systems, many techniques have been developed to represent human knowledge. One of the earliest is to represent knowledge as a logical system of if-then style rules (rule-based systems [5, 13]). A more recent approach is to represent knowledge (including uncertainty) of a situation, or "domain," as a network of states and probabilities (Bayesian Networks [27]).

Many domains, whether they are rule-based, probabilistic, or other, require a representation of time and of the temporal relationships between events. Most systems rely on a mechanism in which a date is associated with each piece of knowledge. Relationships are then determined simply by the date ordering. In more complicated domains, such as emergency room diagnosis, the date mechanism is not sufficient; one must be able to represent situations with relative knowledge like "precedes" or "during."

Real-world domains requiring a unified model of time and uncertainty include dealing with real-time system diagnosis, story understanding, planning and scheduling as well as financial forecasting. For example, consider the following scenario found in computer security analysis:

> The Air Force computer operations center has a secure vault with a time-coded lock. This time-lock allows the vault to be opened from 0900 hours to 0905 hours and from 2100 to 2105. The center has critical operations from 0855 to 1805. Access to the vault is needed during the day and during critical operation making the vault likely to be open at those times. However, if the vault is closed, it can not be reopened until the time-lock allows.

This provides a detailed description of the causal and temporal relationships necessary to properly model the secure vault. As part of the computer security analysis, we must be able to translate this description and capture the knowledge in a form which we can correctly process and reason over.

Once the knowledge representation is captured, inferences can be made. Inferences can be of several types including prediction and explanation. Prediction is concerned with extending forward from the

known past and present to the unknown future (statistical syllogism [16]). Explanation involves the determination of causality by extending from known data back to hypotheses (abduction) [16].

Complex systems consist of collections of interacting processes. These processes change over time in response to both internal and external stimuli as well as to the passage of time itself. There is great variety in the behavior of processes. Some processes are simple events such as opening a door or flipping a switch. Others are complex. One example being a communication channel, in which errors may occur due to lightning strikes and are more likely to occur given previous errors. Processes can also be recurrent or periodic, such as the passing of day into night or shifts in a work schedule.

The problem is to develop a model capable of representing complex systems changing over time. Given evidence about the past and present state of a system, one must be able to predict the system's future state. Also, given a future state, one must be able to determine the most probable causes. As knowledge about such systems is bound to be incomplete and as the systems themselves may not be deterministic, the model must be able to represent uncertainty. This uncertainty permeates all areas, the duration of events, the strength of causal influence, the precise temporal relationship between events, etc. In traditional approaches such as temporal interval algebras and its variations [2, 4, 38, 12, 10, 22], temporal uncertainty is modeled only as a disjunction of the possible temporal relationships between events. For example, event A either occurs before OR during event B. Thus, the goal is to determine a feasible set of relationships between the events that satisfies all the disjunctions. The main limitation with this approach arises when a preference ordering is needed among the relationships in each disjunction as found in the security scenario above.

Bayesian networks [27] provide a robust, probabilistic method of reasoning with uncertainty. Bayesian networks, however, do not provide a direct mechanism for representing temporal dependencies. For example, it is difficult to represent a situation such as the variability of an employee's arrival at work and the causal relationships between the time of arrival and later events.

Prior temporal modeling techniques have made trade-offs in expressiveness between semantics for time and semantics for uncertainty [9]. Significant research has been done exploring time nets (also called time-slice Bayesian networks) [19, 17, 14, 18]. These approaches build on the strong probabilistic semantics of Bayesian networks for expressing uncertainty. The discrete time net approach developed by Kanazawa models time as a series of points [17]. Events are considered to occur at an instant of time while facts are considered to occur over a series of time points. Both events and facts are represented by random variables. If dependencies only connect between random variables at the same or consecutive time points; then the net is said to be a Markov time net. In other words, the Markov property holds for a model when the future is conditionally independent of the past, given the present [19].

Hanks et al [14] is especially interesting for our work due to the emphasis on both endogenous and exogenous change [14]. Endogenous change is triggered by internal action, such as the progression of disease, and exogenous change is triggered by external change such as the administration of drugs. In our model, individual processes within a system must be able to respond to both endogenous (internal) and exogenous (external) stimuli.

The time-sliced approaches mentioned above are based on point models of time and, as such, require that events occur instantaneously. Often it is more natural to consider events as taking place over intervals of time. Also, the relationships between events that occur over intervals can be quite difficult to represent with only the three point relations (precedes, follows, equals).

Santos' Temporal Abduction Problem (TAP) [31] uses an interval representation of time. In the TAP, each event has an associated interval during which the event occurs. Relationships between events are expressed as directed edges from antecedents to consequents within a weighted and/or directed acyclic graph structure. Edges are qualified with the possible interval relations. This allows great flexibility in expressing the relationship between events. For example, say event A is a possible antecedent for an OR-node, event B. For A to be the antecedent supporting B, we might have the additional restriction

that A must occur either before or after event B. The TAP is an extension of cost based abduction [7] using a numeric cost to indicate the uncertainty of an event's occurrence. These costs are generally determined in an ad hoc manner by the domain expert. The TAP trades strong semantics of uncertainty for a powerful and flexible temporal representation.

This paper presents a new model, the Probabilistic Temporal Network (PTN), for representing temporal and atemporal information while remaining fully probabilistic. The model allows representation of time constrained causality, of when and if events occur, and of the periodic and recurrent nature of processes. Bayesian networks lie at the foundation of the system and provide the probabilistic basis. Allen's interval system [2] and his thirteen relations provide the temporal basis.

PTNs focus on directly modeling processes and the interaction between them. The state of a process is represented by a value at a given time interval. A process can be defined over any number of such intervals. Random variables from traditional probability theory are used to model a process' value over each time interval.

We first briefly discuss temporal reasoning and Bayesian networks. From this foundation, the theoretical structure of our model is developed. A linear constraint system for performing belief revision is also developed. Along the way, several examples are developed including the secure vault scenario introduced above.

2 Temporal Reasoning

Temporal reasoning has been defined as the ability to reason about the relationships in time between events [13]. It is necessary to reason about time in many domains including planning, simulation, natural language understanding, and diagnosis. Temporal reasoning has been considered in philosophy and logic since Thales and Zeno [24]; however, it is only in the last two decades that temporal reasoning has been explicitly considered in artificial intelligence. McDermott and Allen, with their work in the

early eighties [2, 3, 4, 25], brought temporal reasoning into the AI mainstream. Other models for temporal reasoning include point algebras [38], semi-intervals [12], temporal constraint networks [10], and weak representations of interval algebras [22]. Much research has also been conducted on the efficient of temporal reasoning [20].

McDermott provides one of the earliest temporal representations [25]. In his approach, time is divided into a series of states with each state having an associated date, i.e. point in time. Facts are expressed as being true during particular states.

Allen introduced interval temporal reasoning to the AI community [2, 4]. Allen's interval algebra is governed by 13 relations on the intervals. Each event has an associated interval, denoted $[a, b]$, where a is the starting time point and b is the termination point. Temporal relationships between events are expressed as relations between their intervals. The relations between intervals, denoted $\mathcal{A}$, are $\{=, <, >, m, mi, d, di, s, si, f, fi, o, oi\}$ [2] (see Table 1). For example, event $A = [a, b]$ preceding event $B = [c, d]$ is denoted $A < B$ indicating that $a < b < c < d$. These relations are mutually exclusive and exhaustive. Note that, while there are thirteen relations between intervals, only three relations exist between points: precedes, equals, and follows.

Symbol:		Name:	Relation:
=	=	equals	
<	>	precedes	
m	mi	meets	
d	di	during	
s	si	starts	
f	fi	finishes	
o	oi	overlaps	

Table 1: The thirteen possible interval-interval relations.

Of special importance is Allen's use of disjunctive sets to express uncertainty in the exact relationship between intervals. For example, "interval A precedes or meets interval B" is written as $A\{<, m\}B$. Some commonly used disjunctions are *disjoint*, written $\{<, >, m, mi\}$, and *contains*, written $\{di, si, fi\}$ [2]. This can be represented in a graphical form where nodes represent events and the arcs are labeled with a disjunction of relations. The goal is to determine whether there exists an interval assignment to all the events that satisfy the disjunctive relations. If such a solution exists, then the given knowledge base is consistent.

While there is debate, in both philosophy and artificial intelligence, as to which representation, points or intervals, is most appropriate; the expressive power of the two methods is generally considered equivalent [17, 2] as intervals can be represented with beginning and end points in a point based approach. Allen points out, however, some paradoxes that can occur when points are allowed as the fundamental unit of time [2]. The problems arise from the durationless nature of points. Durationless intervals are not allowed, i.e. for any interval $[t_1, t_2]$, $t_2 > t_1$. If $t_1 = t_2$ is allowed then the thirteen interval-interval relations are not mutually exclusive. For example $[t_1, t_2]$ *starts* $[t_2, t_3]$ is indistinguishable from $[t_1, t_2]$ *meets* $[t_2, t_3]$ when $t_1 = t_2$. Mathematically, point relations should be expressed as $t_1 \mathcal{R} [t_2, t_3]$ and as such, there is a different set of point-interval relations which would add unnecessary overhead if used in our model. Our model strictly adheres to the philosophy that intervals are primitive and have non-zero duration.

Definition 1. *A* temporal interval *is a closed interval* $[a, b]$ *on the reals (rationals if countability is an issue) such that* $a < b$.

Axiom 1. *The temporal interval is the primitive temporal individual.*

Since all intervals must have non-zero duration, how can point intervals be expressed? The standard approach is to use $[t_0, t_0 + \epsilon]$ where ϵ is arbitrarily close, but not equal, to zero. Note that ϵ can be

either added to the end or subtracted from the beginning or both. This approach is adopted in the PTN. To facilitate specifying the relationships between intervals, ϵ is deemed constant across an entire model. Thus $[t_0, t_0 + \epsilon]\{m\}[t_0, t_1]$ does not hold while $[t_0, t_0 + \epsilon]\{m\}[t_0 + \epsilon, t_1]$ does.

Aside from the temporal domain, neither Allen's nor McDermott's method can explicitly model uncertainty. Uncertainty arises from many sources including missing or unavailable data as well as over generalization of rules [13]. For example if we have the rule "Birds Fly" and "Ostriches are birds" we conclude that "Ostriches fly." To prevent such a conclusion, additional rules must be added such as "Some birds fly" or "Ostriches don't fly." These additional rules can add significant complexity to a knowledge base.

3 Bayesian Networks (BNs)

Approaches to dealing with uncertainty include fuzzy logic [40], cost based techniques [7], certainty factors [34, 35], Dempster-Shafer theory [32], and probabilistic methods [27]. These approaches can be used both extensionally and intensionally. Extensional systems, such as rule-based systems, attach some sort of truth value to each rule or formula. The truth-value for some formula is calculated functionally from the truth-value of sub-formulae. Intensional systems, such as model-based systems, attach uncertainty to the possible states of the system itself [27]. Extensional systems are generally computationally efficient but their uncertainty measures are semantically weak. Intensional systems, on the other hand, are generally computationally expensive and semantically strong [27]. By carefully restricting which parts of an intensional system are relevant to the other parts, the computational limitations can, to some degree, be overcome.

In probabilistic reasoning, random variables (RVs) are used to represent events and/or objects in the world. By assigning various values to these RVs, we can model the current state of the world and weight the states according to the joint probabilities.

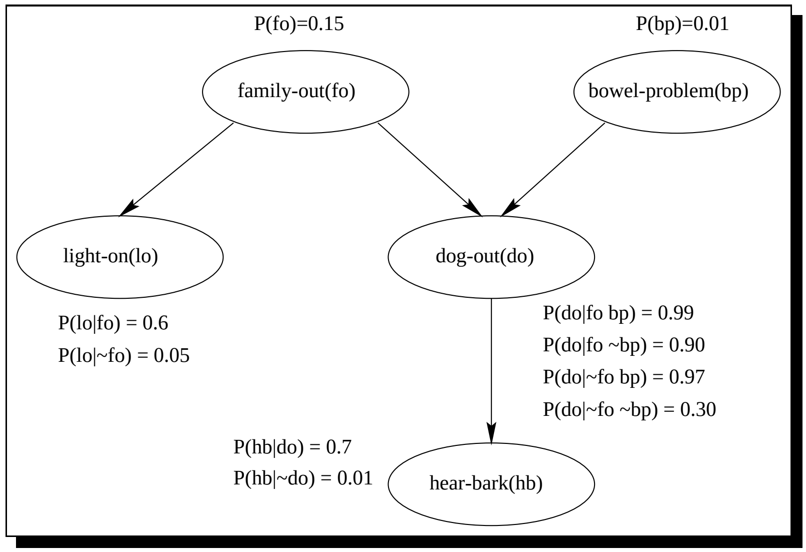

Figure 1: "Suppose when I go home at night, I want to know if my family is home before I try the doors. Now often when my wife leaves the house, she turns on an outdoor light. However, she sometimes turns on this light if she is expecting a guest. Also, we have a dog. When nobody is home, the dog is put in the back yard. The same is true if the dog has bowel troubles. Finally, if the dog is in the backyard, I will probably hear her barking." [6]

Bayesian networks are probabilistic intensional systems in which independence assumptions are used to restrict relevance. A Bayesian network is a directed acyclic graph (DAG) of random variable (RV) relationships. Directed arcs between RVs represent conditional dependencies. When all the parents of a given RV are instantiated, that RV is said to be conditionally independent of the remaining, non-descendent RVs given knowledge of its parents. For a more formal description of the independence semantics in Bayesian networks, see *d-separation* and *I-maps* in Charniak [6] and Pearl [27]. Figure 1 presents a simple example of a Bayesian network.

In general, we are searching for the world state with highest likelihood. This is called *belief revision* [27]. Belief revision is best used for modeling explanatory/diagnostic tasks. Basically, some evidence or observation is given to us, and our task is to come up with a set of hypotheses that together constitute the most satisfactory explanation/interpretation of the evidence at hand. Belief revision is a form of *abductive reasoning* [15, 28, 7]. More formally, if W is the set of all RVs in our given Bayesian network

and e is our given evidence[1], any complete instantiation to all the RVs in W that is consistent with e is called an *explanation* or *interpretation* of e. The problem, then, is to find an explanation w^* such that

$$P(w^*|e) = \max_w P(w|e). \tag{1}$$

w^* is known as the *most-probable explanation*. The joint probability of any explanation w,

$$w = (X_1 = x_1) \wedge (X_2 = x_2) \wedge \ldots \wedge (X_m = x_m) \tag{2}$$

(where $X_1 \ldots X_i \ldots X_m$ is an arbitrary ordering of random variables in W, and x_i is some assignment to random variable X_i) is found using the chain rule [27]:

$$P(w) = P(x_m|x_{m-1}, \ldots, x_1) \cdot P(x_{m-1}|x_{m-2}, \ldots, x_1) \cdots P(x_2|x_1) \cdot P(x_1) \tag{3}$$

Bayesian networks take the chain rule one step further by making the important observation that certain RV pairs may become uncorrelated once information concerning other RV(s) is known. More precisely, we may have the following independence condition:

$$P(A|X_1, \ldots, X_n, U) = P(A|X_1, \ldots, X_n) \tag{4}$$

for some collection of RVs U. Intuitively, we can interpret this as saying that given knowledge of $X_1, \ldots, X_n$ knowledge of U is irrelevant to the state of A.

Combined with the chain rule, these conditional independencies allow us to replace the terms in the chain rule with smaller conditionals. Thus, instead of explicitly keeping the joint probabilities,

[1] That is, e represents a set of instantiations made on a subset of W.

all we need are smaller conditional probability tables, from which the joint probabilities can then be calculated.

For example, an application of the chain rule for computing the probability of an explanation for the Bayesian network in Figure 1 is

$$P(hb, do, lo, fo, bp) = P(hb|do, lo, fo, bp) \cdot P(do|lo, fo, bp) \cdot P(lo|fo, bp) \cdot P(fo|bp) \cdot P(bp) \tag{5}$$

Using the dependencies we can simplify this to

$$P(hb, do, lo, fo, bp) = P(hb|do) \cdot P(do|fo, bp) \cdot P(lo|fo) \cdot P(fo) \cdot P(bp) \tag{6}$$

By choosing an ordering of the random variables consistent with the structure of the graph, such as that used in Equation 6 above, the savings from independencies is maximal and computation from the dependency tables in the Bayesian network is straightforward.

As to the source of the conditional probabilities in a Bayesian network, the values associated with each node only attain meaning after the inference engine reasons over them during belief updating. It should be obvious, however, the inference engine's propagation of probabilities must begin somewhere. In his discussion of the validity of such values to probabilistic reasoning schemes, Pearl [27] writes:

> [p. 148, The] conditional probabilities characterizing the links in the network do not seem to impose definitive constraints on the probabilities that can be assigned to the nodes. ... The result is that any arbitrary assignment of beliefs to the propositions a and b can be consistent with the value of P$(a|b)$ that was initially assigned to the link connecting them ....

Bayesian networks [27] are a natural method for representing uncertainty. Bayesian networks, how-

ever, do not provide a direct mechanism for representing temporal dependencies. For example, it is difficult to represent a situation such as the variability of an employee's arrival at work and the causal relationships between the time of arrival and later events.[2]

4 Combining Time and Probability

As previously discussed, the time-sliced approaches provide strong probabilistic semantics for representing uncertainty, however they are constrained in their temporal expressiveness. The TAP, on the other hand, has strong interval-based temporal semantics, but lacks strong probabilistic semantics.

What is needed, then, is a combined approach integrating strong probabilistic and temporal semantics. While much research has been done on point-based probabilistic temporal network models, little or no research has been identified using interval methods, specifically Allen's interval relations, for intensional probabilistic reasoning. As mentioned earlier, the interval representation of time is important for the expressive set of relations available. The closest research is the temporal abduction problem discussed above which does not have strict probabilistic semantics. Recent work by Young and Santos [39][3] does present a starting point, defining the network structure for a new model. The nodes of the network are what we call *temporal aggregates (TAs)* and the edges are the causal/temporal relationships between aggregates. Each aggregate represents a process changing over time. A temporal aggregate contains every interval of interest for the process. Each interval has an associated random variable giving the state of the process over that interval. For example, to model the state of the secure value in our computer security example earlier, we first identify the time intervals that we are interested in: $[0000 - 0900]$, $[0900 - 1200]$, $[1200 - 2100]$, and $[2100 - 2400]$. We associate a RV with each interval say $o1$, $o2$, $o3$, and $o4$ each with state assignments $\{\mathsf{true}, \mathsf{false}\}$ such that the vault is open during the

[2] For discussions on causality and correlations, see [21, 27, 23, 11].

[3] In which Probabilistic Temporal Networks (PTNs) are termed Temporal Bayesian Networks (TBNs) and Temporal Aggregates (TAs) are termed Temporal Random Variables (TRVs)

interval if its associated RV is true. Since the vault being open is also dependent upon whether critical operations (CO) are in progress and/or the time-unlock (TU) mechanism is in effect, each RV $o1$, $o2$, $o3$, and $o4$ is dependent on these two states. For example, if the vault was closed in the previous interval and the time-unlock mechanism is not activated, then the vault cannot be open in the next interval. In Figure 2, we depict this situation by the fact that $P(ox|\neg VO, CO, \neg TU) = 0$ where ox can be substituted by $o1$, $o2$, $o3$, or $o4$. Thus, this is an aggregate modeling when a vault is open. The 'Vault-Open' TA is dependent on itself (VO) and two other processes (TU and CO). The probabilities, in this case, can be based on frequency. This example is expanded into a full network for the complete scenario later.

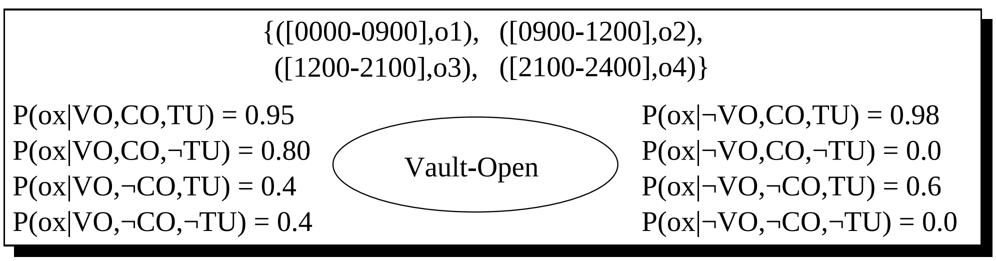

Figure 2: A simple temporal aggregate, 'Vault-Open', defined over four intervals. The conditional probability tables show 'Vault-Open' to be dependent on itself through some temporal causal relationship. ox is simply a place holder and is meant to be substituted by $o1$, $o2$, $o3$, and $o4$.

As is the case in the real world, the apparent state of a process is dependent on the temporal perspective of observation. An observation made in the middle of the night as to whether or not someone is at work may return different results than if the observation is made during the day. A switch can be turned on only if, at some previous time, the switch was turned off; the light can be on only when the switch is on.

To model the difference perspective makes in the apparent state of a process, edges in the network consist of a disjunctive set of interval relations and a schema to map the random variables of the intervals to a single value. This allows the exact definition of those intervals during which the state of one process affects another.

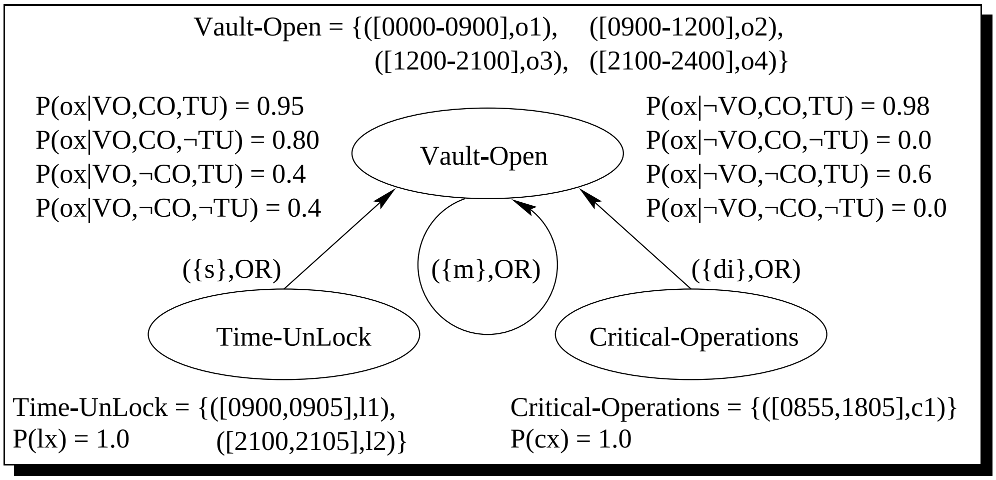

Figure 3: A probabilistic temporal network modeling a secure vault. This extends the 'Vault-Open' temporal aggregate in Figure 2.

Figure 3 shows a probabilistic temporal network modeling our earlier secure vault scenario detailing the various components and their interactions.

4.1 Temporal Aggregates

A process, such as 'Vault-Open' in Figure 3, is represented in the PTN by a temporal aggregate. Intuitively, a temporal aggregate consists of the set of states, e.g. $\{\mathsf{true}, \mathsf{false}\}$, $\{1, 2, 3\}$, or $\{\mathsf{false}\} \cup \{Red, Blue\}$, that the process can take on, and a set of temporal intervals each having an associated random variable. Each such RV has a conditional probability table defined over the states of the process.

Definition 2. *A* temporal aggregate *(TA) is an ordered pair* (T, Σ) *in which* Σ *is a set of states and* T *(pronounced* Tau*) is a set of ordered pairs* (i, r) *where* i *is a temporal interval and* r *is a random variable defined over* Σ*. For all pairs* (i_1, r_1) *and* (i_2, r_2) *in* T, $r_1 = r_2$ *iff* $i_1 = i_2$*. The dependencies for each random variable in the TA are defined only by temporal causal relationships between TAs.*

In our prior work [39], temporal aggregates (then termed TRVs) were allowed to have internal dependencies to model endogenous change. This was found to be a source of temporal inconsistency and better represented through self loops as demonstrated in Figure 3. Endogenous change is explicitly

modeled in the PTN with cyclic temporal causal relationships. This can be seen in the 'Vault-Open' process in Figure 3 in which the vault is more likely to stay open, given that it is open. Also note that this definition allows T to contain a potentially infinite number of interval-RV pairs. This paper assumes that temporal aggregates are finite, both in T and in Σ.

'Vault-Open' is formally written, according to Definition 2, as $VO = \{\mathrm{T}, \Sigma\}$ where

$$\mathrm{T}_{VO} = \{([0000, 0900], o_1), ([0900, 1200], o_2), ([1200, 2100], o_3), ([2100, 2400], o_4)\}$$

and

$$\Sigma_{VO} = \{\mathsf{true}, \mathsf{false}\}$$

with the conditional probability table being

$$\begin{array}{lclllcl} P(o_x|VO, CO, TU) & = & 0.95 & & P(o_x|\neg VO, CO, TU) & = & 0.80 \\ P(o_x|VO, CO, \neg TU) & = & 0.80 & & P(o_x|\neg VO, CO, \neg TU) & = & 0.0 \\ P(o_x|VO, \neg CO, TU) & = & 0.4 & & P(o_x|\neg VO, \neg CO, TU) & = & 0.6 \\ P(o_x|VO, \neg CO, \neg TU) & = & 0.4 & & P(o_x|\neg VO, \neg CO, \neg TU) & = & 0.0 \end{array}$$

for all RVs o_x where $o_x \in \{o_1, o_2, o_3, o_4\}$. The $\neg$ symbol (as in $\neg TU$ above) indicates that the RV is assigned false, a non-negated RV (as in TU) indicates that the RV is assigned true.

Since $\Sigma = \{\mathsf{true}, \mathsf{false}\}$, $P(\neg o_x|VO, CO, TU) = 1 - P(o_x|VO, CO, TU)$. This holds for the other probabilities as well. In general, we will not explicitly show the probabilities when the true case is zero; e.g. $P(o_x|\neg VO, \neg CO, \neg TU) = 0.0$ would not be shown. Symbols used for temporal aggregates are uppercase letters from the end of the alphabet, e.g. X or Y, or uppercase abbreviations from the text name of the process being modeled, e.g. process 'Vault-Open' has a temporal aggregate denoted VO. Random variables within temporal aggregates are denoted with lowercase letters, e.g. a, b, and c or

y_1 and y_2. Since the possible states of the aggregate are often evident from the conditional probability tables, Σ is often not explicitly shown. To differentiate between components of different temporal aggregates, the symbol of the component can contain the subscripted symbol of the associated TA, e.g. Σ_{VO} or $o_{1_{VO}}$.

An assignment to a temporal aggregate consists of an assignment to each interval-RV pair.

Definition 3. *A is an* aggregate assignment *(AA) iff A is a set of ordered pairs (τ, σ) where $\tau \in \mathrm{T}$ and $\sigma \in \Sigma$ such that $\forall \tau \in \mathrm{T}$, there exists an unique $\sigma \in \Sigma$ such that $(\tau, \sigma) \in A$. In other words, an* aggregate assignment *is a function from* T *into* Σ.

Intuitively, an aggregate assignment for a single temporal aggregate is simply assigning a state to each RV involved with the intervals in the aggregate. For example,

$$A_{VO} = \{([0000, 0900], \mathsf{false}), ([0900, 1200], \mathsf{true}), ([1200, 2100], \mathsf{true}), ([2100, 2400], \mathsf{false})\}$$

is an AA for the temporal aggregate VO from Figure 2. A_{VO} might be read "The vault was closed from 0000 hours to 0900 hours, open from 0900 hours to 2100 hours, and closed from 2100 hours to 2400 hours." The use of past tense here is arbitrary, *is closed* or *will be closed* would be equally appropriate. Aggregate assignments are denoted by uppercase letters from the beginning of the alphabet, e.g. A or B, subscripted if necessary by the symbol for the associated temporal aggregate.

Sometimes the entire state of a TA is not known. For example, we may only know that the vault was closed from 0000 to 0900. To express this we use a partial aggregate assignment which is simply a subset of an aggregate assignment.

Definition 4. *P is a* partial aggregate assignment *(PAA) for some temporal aggregate, X, iff there exists an A such that $P \subseteq A$ where A is an aggregate assignment for X. In other words, a partial aggregate assignment is a partial function from* T *into* Σ*.*

Our example, where the vault is only known to be closed over one interval is thus written:

$$P_{VO} = \{([0000, 0900], \textsf{false})\}$$

Note that P_{VO} is a subset of aggregate assignment A_{VO} above. PAAs are usually denoted by capital letters from the middle of the alphabet; however, since, by definition an aggregate assignment is also a PAA, some uppercase letters from the beginning of the alphabet may sometimes be used for PAAs.

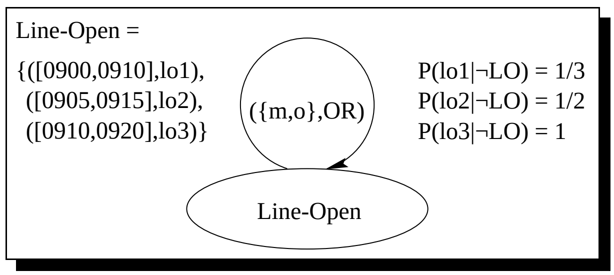

Figure 4: A simple, one process probabilistic temporal network enforcing a mutual exclusion relationship. A communication line can only be opened given that it has not previously been opened.

4.2 Temporal Causal Relationships

How are the aggregates interconnected? The example network in Figure 4 shows a directed edge from 'Line-Open' to itself labeled $(\{m, o\}, \mathbf{OR})$. The edge combined with the conditional probability tables enforce a mutual exclusion constraint on 'Line-Open'. The communication line can be opened only if the line was not previously opened. Edges in the probabilistic temporal network are temporal causal relationships or TCRs.

While portrayed graphically as a labeled edge between temporal aggregates, the TCR is actually shorthand for a set of induced random variables that enforce the temporal constraints. These random variables combine the intervals selected by a disjunctive set of interval relations, e.g. $\{o, m\}$, using the probability distribution specified by a schema, e.g. **OR**, **XOR**, **PASSTHROUGH**. Figure 5 shows our example network from Figure 4 with the intervals and temporal relations explicitly shown. For

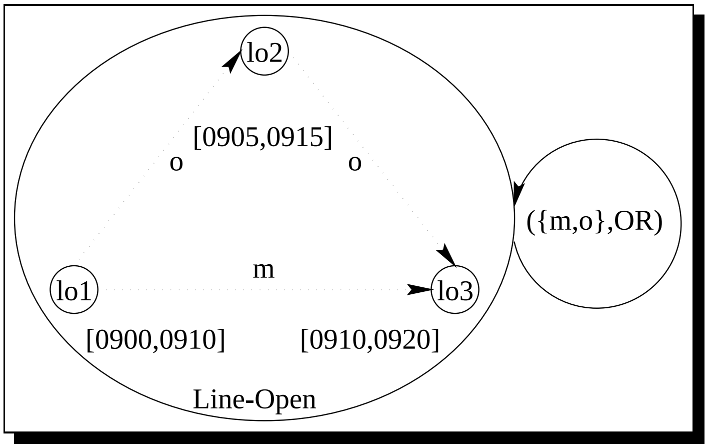

Figure 5: The probabilistic temporal network from Figure 4 broken out to explicitly show the intervals (small circles) and the temporal relationships between intervals (dotted lines).

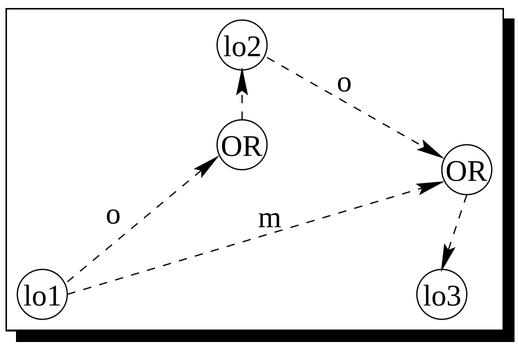

Figure 6: The network in Figure 5 with the temporal causal relation replaced with the TCR induced random variables.

example, the dotted line from interval lo_1 to interval lo_2 shows that lo_1 overlaps lo_2. Figure 6 shows the network with the TCR replaced by the appropriate induced RVs.

What are the semantics behind the temporal causal relationship? The probability of some TA Y taking on some particular state over each interval is dependent on TA X taking on some state on interval(s) fitting the temporal relation, e.g. "no interval in Y can have state `true` unless that interval is after some interval in X having state `true`." This is written $X(\{<\}, \mathbf{OR})Y$ with every $(i, r) \in \mathrm{T}(Y)$ having conditional probabilities of the form $P(r|\ldots, \neg X) = 0.0$. Schemas in general and the **OR** schema in particular are further discussed below.

Definition 5. *A* temporal causal relationship *(TCR) describes a relationship between two temporal*

aggregates $X = (\mathrm{T}_X, \Sigma_X)$ *and* $Y = (\mathrm{T}_Y, \Sigma_Y)$ *where* X *is considered the "cause" and* Y *the "effect." Textually, the TCR is written* $X(\mathcal{R}, \mathcal{M})Y$ *where* $\mathcal{R}$ *is a nonempty set of interval relations and* $\mathcal{M}$ *is a schema for describing random variables. Graphically, the TCR is presented as a directed edge from the node for* X *to the node for* Y*, labeled with* $(\mathcal{R}, \mathcal{M})$*. Formally, the relationship is written as the four-tuple* $(\mathcal{R}, \mathcal{M}, X, Y)$*.*

The TCR induces, for each interval-RV pair, (i_Y, r_Y) *in* T_Y*, a random variable* $\mathcal{M}_r$*, defined over* Σ_X*, such that*

1. r_Y *is directly dependent on* $\mathcal{M}_r$*.*

2. *for each* $(i_X, r_X) \in \mathrm{T}_X$ *where* $i_X \mathcal{R} i_Y$*,* $\mathcal{M}_r$ *is directly dependent on* r_X*.*

3. *for each random variable* x *such that* $\mathcal{M}_r$ *is directly dependent* x*, there exists an* i_X *such that* $(i_X, x) \in \mathrm{T}_X$*.*

4. *the conditional probability table for* $\mathcal{M}_r$ *is defined by the schema* $\mathcal{M}$*.*

Temporal causal relationships are rarely given explicit names. Notationally, the random variables in the interval-RV pairs in the effect TA are usually written, in the conditional probability tables, as being dependent simply on the cause TA. This can be seen in the tables for the 'Vault-Open' temporal aggregate in Figure 3. In cases where there is more than one TCR between two TAs, some appropriate name or symbol can be associated with the TCR and the dependencies in the effect TA can be written as the name of the cause TA subscripted with the name of the TCR.

The random variable schema algorithmically defines the conditional probability tables for the random variables induced by the temporal causal relationship.

Definition 6. *A* random variable schema $\mathcal{M}$ *takes as parameters a set of states* Σ*, a set of interval-RV pairs* T *with RVs defined over* Σ*, a single interval-RV pair* (i, r)*, and an algorithm* Δ *which together*

define the conditional probability table for a random variable $\mathcal{M}_r$ with states Σ such that for each $(i_{\mathrm{T}}, r_{\mathrm{T}}) \in \mathrm{T}$, $\mathcal{M}_r$ is directly dependent on r_{T}. $\mathcal{M}_r$ is directly dependent on nothing else. The conditional probability table for $\mathcal{M}_r$ is constructed with an algorithm, Δ. Δ can be either declarative or procedural.

These schemas can be extremely simple, e.g.

$$\mathbf{OR} : (\mathrm{T}, \Sigma = \{\mathsf{true}, \mathsf{false}\}, (i, r), \Delta_{\mathbf{OR}}) \rightarrow \mathbf{OR}_r$$

where $\Delta_{\mathbf{OR}}$ is defined as

Algorithm 1: ($\Delta_{\mathbf{OR}}$)

1. *Let $(i_{\mathrm{T}_1}, r_{\mathrm{T}_1}) \ldots (i_{\mathrm{T}_n}, r_{\mathrm{T}_n})$ be an arbitrary ordering of the elements of* T
2. *Create random variable $\mathbf{OR}_r$ such that for each assignment A to $\{r_{\mathrm{T}_1}, \ldots, r_{\mathrm{T}_n}\}$*
   (a) *If there exists an $r \in A$ such that $r = \mathsf{true}$*

   Let $$\begin{aligned} P(\mathbf{OR}_r = \mathsf{true}|A) &= 1 \\ P(\mathbf{OR}_r = \mathsf{false}|A) &= 0 \end{aligned}$$

   (b) *else*

   Let $$\begin{aligned} P(\mathbf{OR}_r = \mathsf{true}|A) &= 0 \\ P(\mathbf{OR}_r = \mathsf{false}|A) &= 1 \end{aligned}$$

Exclusive-or, **XOR**, can be defined by changing "there exists an $r \in A$" in step 2a above to "there exists a unique $r \in A$." The other logical operations are also easily defined.

The schema **PASSTHROUGH**, defined:

$$\mathbf{PASSTHROUGH} : (\mathrm{T} = (i_{\mathrm{T}}, r_{\mathrm{T}}), \Sigma, (i, r), \Delta_{\mathbf{PASSTHROUGH}}) \rightarrow \mathbf{PASSTHROUGH}_r$$

with $\Delta_{\mathbf{PASSTHROUGH}}$ defined as

Algorithm 2: ($\Delta_{\text{PASSTHROUGH}}$)

1. *Create random variable* $\mathbf{PASSTHROUGH}_r$ *such that for each* $\sigma \in \Sigma$

$$
\begin{aligned}
P(\mathbf{PASSTHROUGH}_r = \sigma | r_{\mathrm{T}} = \sigma) &= 1 \\
P(\mathbf{PASSTHROUGH}_r \neq \sigma | r_{\mathrm{T}} = \sigma) &= 0
\end{aligned}
$$

produces a random variable for a causal relationship from a singleton TA (only one interval-RV pair in T). The temporal causal relationship $X(\mathcal{A}, \mathbf{PASSTHROUGH})Y$, read "$X$ exerts direct causal influence on Y under all temporal relationships" is analogous to the a non-temporal relation in Bayesian networks. This type of relationship is useful when 'temporalizing' existing Bayesian networks.

4.3 Probabilistic Temporal Networks

A probabilistic temporal network is a directed graph in which the nodes are TAs and the edges are temporal causal relationships.

Definition 7. *A* probabilistic temporal network *(PTN) is an ordered pair* (R, E) *where* R *is a set of temporal aggregates and* E *is set of temporal causal relationships such that, for each TCR in* E *from some temporal aggregate,* X*, to some temporal aggregate,* Y*, both* X *and* Y *are in* R*.*

If each temporal aggregate in a probabilistic temporal network is assigned, then that PTN is said to be completely assigned. The set of all of the assignments and associated temporal aggregates forms a complete assignment.

Definition 8. *The set* $\mathscr{C}$ *containing (temporal aggregate, aggregate assignment) pairs is a* complete assignment *(CA) of some PTN* (R, E) *iff*

1. $\forall (X, A) \in \mathscr{C}$, $X \in R$ *and* A *is an aggregate assignment of* X*.*

2. $\forall (X, A), (Y, B) \in \mathscr{C}$, $X = Y \Rightarrow A = B$*.*

3. $\forall X \in R \exists (Y, A) \in \mathscr{C}$ *such that* $X = Y$.

Complete assignments are denoted by uppercase script letters from the beginning of the alphabet, e.g. $\mathscr{A}$,$\mathscr{B}$, or $\mathscr{C}$.

When inferencing over a probabilistic temporal network, incomplete evidence as to the state of the network may be held. Such evidence is represented with a partial assignment. In the simplest form, any subset of a complete assignment is a partial assignment. A more complicated case arises when only a partial aggregate assignment is known for some temporal aggregate. Since a PAA is a subset (possibly improper) of an aggregate assignment, a partial assignment to a PTN consists of a subset of the variables of the PTN and associated partial aggregate assignments for the TAs. More formally:

Definition 9. *The set* $\mathscr{P}$ *containing (temporal aggregate, aggregate assignment) pairs is a* partial assignment *(PA) of some PTN* (R, E) *iff*

1. $\forall (X, P) \in \mathscr{P}$, $X \in R$ *and* P *is a partial aggregate assignment of* X.

2. $\forall (X, P), (Y, Q) \in \mathscr{P}$, $X = Y \Rightarrow P = Q$.

PAs are usually denoted with uppercase script letters from the middle of the alphabet, e.g. $\mathscr{P}$ or $\mathscr{Q}$. As a complete assignment is a subset of itself, by definition any complete assignment is also a partial assignment.

Notation. *A partial assignment,* $\mathscr{P}$*, is said to be a subset of another partial assignment,* $\mathscr{Q}$*, (denoted* $\mathscr{P} \sqsubseteq \mathscr{Q}$*) if every* (X, P) *in* $\mathscr{P}$ *(except those having* $P = \emptyset$*) has a corresponding* (Y, Q) *in* $\mathscr{Q}$ *such that* $X = Y$ *and* $P \subseteq Q$*. A complete assignment, say* $\mathscr{C}$*, is said to be* compatible *with a partial assignment,* $\mathscr{P}$*, if* $\mathscr{P} \sqsubseteq \mathscr{C}$*, otherwise* $\mathscr{C}$ *is said to be* incompatible *with* $\mathscr{P}$*. If* $\mathscr{C}$ *is incompatible with* $\mathscr{P}$ *then at least one temporal aggregate in* $\mathscr{C}$ *has a different assignment than that in* $\mathscr{P}$*.*

The goal of belief revision is to find the most probable state of the world given some evidence. This is the *most probable explanation.*

Definition 10. *Let B be a PTN, let $\mathscr{P}$ be partial assignment (evidence) of B, and let $\mathscr{C}$ be some complete assignment (explanation) of B. $\mathscr{C}$ is a* most probable explanation *(MPE) given $\mathscr{P}$ iff for all $\mathscr{A}$ where each $\mathscr{A}$ is a complete assignment of B compatible with $\mathscr{P}$, $P(\mathscr{C}|\mathscr{P}) \geq P(\mathscr{A}|\mathscr{P})$.*

Since $P(\mathscr{A}|\mathscr{P}) = P(\mathscr{A}, \mathscr{P})/P(\mathscr{P})$ and an incompatible complete assignment can not be a MPE (unless the evidence $\mathscr{P}$ is itself contradictory in which case all CAs are MPEs), we only need to consider as candidates those complete assignments for which $\mathscr{P} \sqsubseteq \mathcal{A}$. Thus since $\mathscr{P} \sqsubseteq \mathscr{A}$, we derive $P(\mathscr{A}|\mathscr{P}) = P(\mathscr{A})/P(\mathscr{P})$. Furthermore, since $1/P(\mathscr{P})$ is a factor in the conditional probability of each explanation $\mathscr{A}$, to find the MPE, we need only compute the probability of each complete assignment, i.e. $P(\mathscr{A})$. $P(\mathscr{A})$ is calculated with the chain rule.

5 Example Computation of Joint Probability

Previously we discussed finding the *most probable explanation.* The MPE is the complete assignment with the greatest joint probability. As mentioned, this joint probability is calculated using the chain rule.

Joint Probability Table for Figure 4						
Line-Open					**Assignment Probability:**	
[0910,0920]	**OR$_{lo_3}$**	**[0905,0915]**	**OR$_{lo_2}$**	**[0900,0910]**		
true 1	false 1	false 1/2	false 1	false 2/3	1/3	(1)
false 1	true 1	true 1/2	false 1	false 2/3	1/3	(2)
false 1	true 1	false 1	true 1	true 1/3	1/3	(3)
true 0	true 1	false 1	true 1	true 1/3	0	(4)
				Total:	1	

Table 2: The possible complete assignments to the network in Figure 4 with associated probabilities. One 'impossible' assignment is also shown.

Table 2 shows the probability distribution defined by the example in Figure 4. Only non-zero probability assignments are shown (but one). Each joint probability in Table 2 is calculated using the

chain rule [27]. For example, the probability of the complete assignment

$$\{(LO, \{((([0900, 0910], lo_1), \mathsf{true}), (([0905, 0915], lo_2), \mathsf{false}), (([0910, 0920], lo_3), \mathsf{false})\})\} \tag{7}$$

is calculated from

$$P(lo_3 = \mathsf{false}|\mathbf{OR}_{lo_3} = \mathsf{true}) * P(\mathbf{OR}_{lo_3} = \mathsf{true}|lo_1 = \mathsf{true}, lo_2 = \mathsf{false}) *$$

$$P(lo_2 = \mathsf{false}|\mathbf{OR}_{lo_2} = \mathsf{true}) * P(\mathbf{OR}_{lo_2} = \mathsf{true}|lo_1 = \mathsf{true}) *$$

$$P(lo_1 = \mathsf{true}) =$$

$$1.0 * 1.0 * 1.0 * 1.0 * \tfrac{1}{3} = \tfrac{1}{3}$$

6 Cycles and Temporal Ordering

Now that the basic definitions and properties have been introduced, we will briefly explore the probabilistic temporal network in Figure 4 and consider a potential alternate representation. Figure 4 shows a network using a cyclic dependency to represent the internal dependencies in process 'Line-Open', i.e., a cyclic TCR has been used to explicitly model the endogenous temporal relationships. For 'Line-Open' to be true over some interval, 'Line-Open' must not be true over any earlier intervals.

Examining the intervals, "earlier" turns out to be either *meets* or *overlaps*. This is represented with a disjunctive set containing *meets* and *overlaps*: $\{m, o\}$. The conditional dependencies are represented using the **OR** schema. The TCR, $LO(\{m, o\}, \mathbf{OR})LO$, describes the random variable $\mathbf{OR}_{lo_3}$ such that $P(\mathbf{OR}_{lo_3}|\neg lo_1, \neg lo_2) = 0$ and $P(\neg\mathbf{OR}_{lo_3}|\neg lo_1, \neg lo_2) = 1$. $\mathbf{OR}_{lo_3}$ replaces LO in $P(lo_3|\neg LO) = 1$ to yield $P(lo_3|\neg\mathbf{OR}_{lo_3}) = 1$. By using cyclic TCRs to explicitly represent the temporal relationships within a process, the knowledge engineer can more clearly "see" the nature of the system being modeled.

Figure 7 shows an attempt to simplify the conditional dependencies in process 'Line-Open'. The

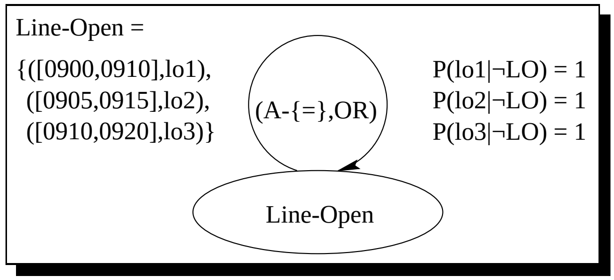

Figure 7: The network in Figure 4 rewritten using a cyclic dependency such that the conditional probability table for each RV can be written with the same probability 1 instead of the dependent probabilities 1/3, 1/2, and 1 (not well-formed).

conditional probability tables for each random variable in process LO are identical. This is accomplished using the TCR $LO(\mathcal{A} - \{=\}, \mathbf{OR})LO$[4], which states that the random variable in each interval-RV pair is dependent on the random variables in all the other interval-RV pairs. While visually similar to the network in Figure 4, there is a serious problem with this network.

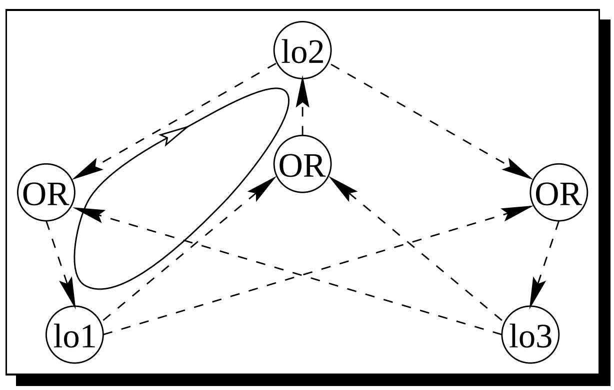

Figure 8: Process 'Line-Open' from Figure 7 drawn with the TCR $LO(\mathcal{A} - \{=\}, \mathbf{OR})LO$ expanded. The loop shows a cycle in the dependencies.

Figure 8 shows process 'Line-Open' with the TCR expanded into its underlying random variable components. Notice that this potentially violates the conditional independence assumptions. Random variable lo_2 is dependent on $\mathbf{OR}_{lo_2}$ which is dependent on lo_1 which is dependent on $\mathbf{OR}_{lo_1}$ which is dependent on lo_2 which is .... lo_2 is separated from itself by random variables $\mathbf{OR}_{lo_2}$, lo_1, and $\mathbf{OR}_{lo_1}$ indicating that given knowledge of each of these variables that lo_2 is independent of itself which is

[4]The set, $\mathcal{A} - \{=\}$, consists of all thirteen interval relations sans *equals*

clearly contradictory.

Figure 4 demonstrates an example in which a cycle in the PTN provided a useful representation of the internal dependencies within a process. Figure 7, on the other hand, shows a case in which the cycle, while intuitively satisfying, violates the requirements of conditional independence. This raises the question: "Under what circumstances are cycles appropriate in probabilistic temporal networks?"

Cyclic dependencies potentially occur when an interval-RV pair becomes self-dependent. If only temporal relations which are strictly one directional are used, an interval-RV pair can not possibly be self-dependent. For example, if only $\{<\}$ is used in a PTN, no cycles are possible. The authors, in the development of the temporal abduction problem, defined the concept of *monotonicity* [30] as applied to temporal relations.

Definition 11. *A set $\mathcal{R}$ of temporal relations is said to be* monotonic *if and only if for all R in $\mathcal{R}$, $\overline{R}^c \cap (\overline{R}^c)^{-1} = \emptyset$ where $\overline{R} = \cup_{R \in \mathcal{R}} R$ and $\overline{R}^c$ is the transitive closure of R and $\overline{R}^{c-1}$ is the inverse of the transitive closure of R.*

In the same work, we introduced the following monotonic set:

Proposition 1. *The subset of relations $\mathcal{C} = \{<, o, s, fi, di, m\}$ from the original thirteen is a monotonic set.*

Intuitively, a monotonic set, such as $\mathcal{C}$ above, can be said to temporally 'point in only one direction.' This is compatible with Suppes' probabilistic theory of causality [36] and Shoham's criteria for causation [33] (both point based approaches) in which causation can only extend forward in time. For this reason, $\mathcal{C}$ is said to be the *causal* set of temporal relations. The network in Figure 4 holds on $\mathcal{C}$.

Theorem 1. *If, for probabilistic temporal network (R, E), there exists a* monotonic *set, $\mathcal{Q}$, of temporal relations such that for each $(\mathcal{R}, \mathcal{M}, X, Y) \in R$, $\mathcal{R} \subseteq \mathcal{Q}$; then the PTN (R, E) is* well-formed.

Proof. Since the only temporal relations used in the PTN are drawn from $\mathcal{Q}$ and $\mathcal{Q}$ is *monotonic*, no interval-RV pair can ever relate to itself temporally (otherwise $\overline{Q}^c \cap (\overline{Q}^c)^{-1} \neq \emptyset$) and as there can be no cycles within the TAs themselves; there can be no cycles in the PTN. □

Combining Theorem 1 and the *causal* set $\mathcal{C}$ from Proposition 1 leads us to the following definition:

Definition 12. *A causal probabilistic temporal network (CPTN) is a PTN for which Theorem 1 holds.*

The causal PTN model enforces the constraint that causality flow forward in time. Each link in the network advances in time. When following a cycle from a temporal aggregate back to itself, one always returns to a different interval-RV pair. The CPTN model enforces, through local constraints, a consistent ontological theory of time.

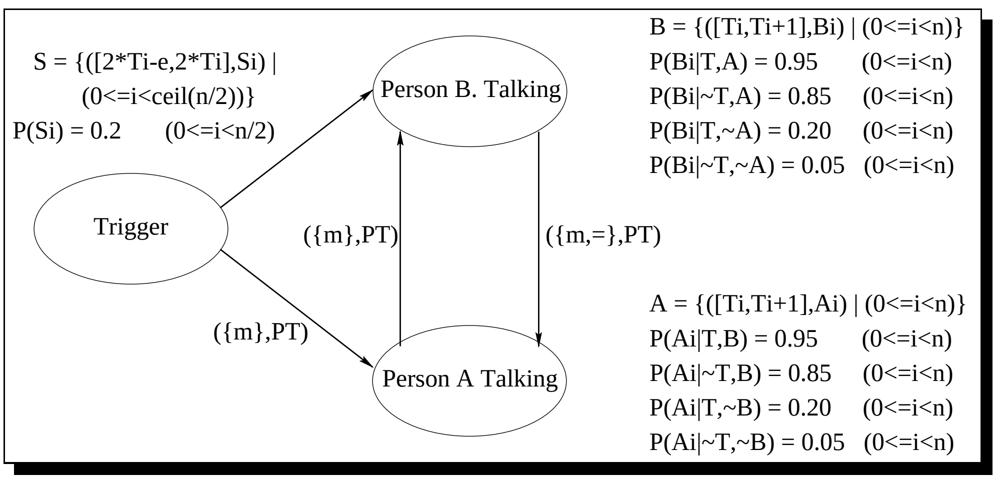

Figure 9: PTN modeling two people chatting with an an occasional conversational trigger. Note the use of set-builder notation.

The equals relation, $=$, is not a member of $\mathcal{C}$, and can not be a member of any *monotonic* set of relations as $=$ is its own inverse. Equals is, however, useful for expressing simultaneity. Figure 9 shows an example in which two people are chatting. Talker A tends to 'talk over' Talker B. To model this, the TCR from B to A includes equals as well as meets.

To extend CPTNs to incorporate $=$s, each directed cycle must have at least one TCR in which equals is not used. This guarantees 'time progression' in each cycle. A probabilistic temporal network limited

to $\mathcal{C} \cup \{=\}$ with this broken cycle property is said to be S-Causal (SCPTN) ('S' for simultaneity).

7 Constraint Satisfaction

In the Section 5, we showed how to calculate the probability of a complete assignment to a probabilistic temporal network. In this section we present a method for finding the most probable complete assignment, i.e., performing belief revision on probabilistic temporal networks. We use a constraint satisfaction approach with mixed Boolean linear programming. Constraint satisfaction has three main advantages; first, constraints can be formed to take advantage of the inherent structure of the PTN; second, very efficient algorithms developed by the operations research community are available; and finally, alternate explanations, e.g. second or third best, can be found using techniques presented in [29].

Definition 13. *A* constraint system *is a 3-tuple* $(?, I, \psi)$ *where* $?$ *is a finite set of variables,* I *is a finite set of linear inequalities based on* $?$*, and* ψ *is a cost function from* $? \times \{\mathsf{true}, \mathsf{false}\}$ *to* $\Re$.

Our probabilistic temporal network model can be considered to have a layered structure. The layers consist of temporal aggregates and temporal causal relationships. For this reason, we present our system of constraints in two parts, those for TCRs and those for TAs. For some well-formed PTN $P = (R, E)$, the following steps produce the constraints, variables, and costs for the temporal causal relationships in E and those for the temporal aggregates in R, i.e. the following steps produce $L(P) = (?, I, \psi)$.

1. For each TCR $(\mathcal{R}, \mathcal{M}, (\mathrm{T}_X, \Sigma_X), (\mathrm{T}_Y, \Sigma_Y))$ in E,

   (a) For each $(i_Y, r_Y) \in \mathrm{T}_Y$ construct variables $\mathcal{M}^{r_Y}_{\sigma_{X_1}} \ldots \mathcal{M}^{r_Y}_{\sigma_{X_n}}$ in $?$ where $\sigma_{X_1} \ldots \sigma_{X_n}$ are states in Σ_X. Set costs for each variable as

$$\psi(\mathcal{M}^{r_Y}_{\sigma_{X_i}}, \mathsf{false}) = \psi(\mathcal{M}^{r_Y}_{\sigma_{X_i}}, \mathsf{true}) = 0. \tag{8}$$

where $1 \leq i \leq n$ and add the following constraint to I:

$$\sum_{i=1}^{n} \mathcal{M}_{\sigma_{X_i}}^{r_Y} = 1. \tag{9}$$

(b) For each $(i_Y, r_Y) \in \mathrm{T}_Y$ and each $\sigma_X \in \Sigma_X$ let $(i_{X_1}, r_{X_1}) \ldots (i_{X_j}, r_{X_j}) \in \mathrm{T}_X$ be those pairs for which $i_{X_h} \mathcal{R} i_Y$ with $1 \leq h \leq j$, then

i. for each conditional probability of the form

$$P(\mathcal{M}_{r_Y} = \sigma_X | r_{X_1} = \sigma_{X_1} \ldots r_{X_j} = \sigma_{X_j})$$

as induced by schema $\mathcal{M}$, construct a variable

$$q[\mathcal{M}_{r_Y} = \sigma_X | r_{X_1} = \sigma_{X_1} \ldots r_{X_j} = \sigma_{X_j}] \tag{10}$$

(denoted q in following steps) in ? such that

A. $\psi(q, \mathsf{false}) = 0$,$\psi(q, \mathsf{true}) = -log(P(\mathcal{M}_{r_Y} = \sigma_X | r_{X_1} = \sigma_{X_1} \ldots r_{X_j} = \sigma_{X_j}))$

B. with the following constraint in I:

$$q \geq \sum_{h=1}^{j} T_{\sigma_{X_h}}^{r_{X_h}} + \mathcal{M}_{\sigma_X}^{r_Y} - j \tag{11}$$

(c) Let $\Upsilon_{\mathcal{M}_{\sigma_X}^{r_Y}}$ be the set of all q constructed in step (1b) for variable $\mathcal{M}_{\sigma_X}^{r_Y}$. For each such variable, add the following constraint to I:

$$\mathcal{M}_{\sigma_X}^{r_Y} = \sum_{q \in \Upsilon_{\mathcal{M}_{\sigma_X}^{r_Y}}} q. \tag{12}$$

2. For each TA $X = (\mathrm{T}_X, \Sigma_X)$ in R

   (a) For each $(i_X, r_X) \in \mathrm{T}_X$ construct variables $\mathcal{T}^{r_X}_{\sigma_{X_1}} \ldots \mathcal{T}^{r_X}_{\sigma_{X_n}}$ in ? where $\sigma_{X_1} \ldots \sigma_{X_n}$ are states in Σ_X. Set costs for each variable as

$$\psi(\mathcal{T}^{r_X}_{\sigma_{X_i}}, \mathsf{false}) = \psi(\mathcal{T}^{r_X}_{\sigma_{X_i}}, \mathsf{true}) = 0. \tag{13}$$

   where $1 \leq i \leq n$ and add the following constraint to I:

$$\sum_{i=1}^{n} \mathcal{T}^{r_X}_{\sigma_{X_i}} = 1. \tag{14}$$

   (b) For each $(i_X, r_X) \in \mathrm{T}_X$ and each $\sigma_X \in \Sigma_X$ let $\mathcal{M}_1 \ldots \mathcal{M}_j$ be those random variables induced by TCRs $(\mathcal{R}_h, \mathcal{M}_h, Y_h, Z_h)$ for which $1 \leq h \leq j$ and $Z_h = X$. Then

      i. for each conditional probability of the form

$$P(r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}),$$

      construct a variable

$$q[r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}] \tag{15}$$

      (denoted q in following steps) in ? such that

         A. $\psi(q, \mathsf{false}) = 0$,$\psi(q, \mathsf{true}) = -log(P(r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}))$

B. and add the following constraint to I:

$$q \geq \sum_{h=1}^{j} \mathcal{M}_{\sigma_{Y_h}}^{r_X} + \mathcal{T}_{\sigma_X}^{r_X} - j \tag{16}$$

(c) Let $\Upsilon_{\mathcal{T}_{\sigma_X}^{r_X}}$ be the set of all q constructed in step (1b) for variable $\mathcal{T}_{\sigma_X}^{r_X}$. For each such variable, add the following constraint to I:

$$\mathcal{T}_{\sigma_X}^{r_X} = \sum_{q \in \Upsilon_{\mathcal{T}_{\sigma_X}^{r_X}}} q. \tag{17}$$

In this construction, constraints (9) and (14) ensures that each random variable, either induced or in a TA, can take on one and only one value. Constraints (11) and (12) guarantee that each of the probabilities for TCR induced variables is computed in concordance with the appropriate temporal relations and schema. Constraints (16) and (17) guarantee that the probability of a temporal assignment to a TA is computed with the appropriate set of conditional probabilities. Variables of the form $q[r_X = \sigma_X | \mathcal{M}_1 = \sigma_{Y_1} \ldots \mathcal{M}_j = \sigma_{Y_j}]$ are called *conditional variables* in that they explicitly represent the dependencies between RVs and are the mechanism for computing the probability of any complete assignment.

For example, consider again the simple probabilistic temporal network in Figure 4. Previously we demonstrated how to calculate the probability of an assignment to this network using the chain rule (see Table 2). Now, if we take the complete assignment

$$\{(LO, \{(([0900, 0910], lo_1), \mathsf{true}), (([0905, 0915], lo_2), \mathsf{false}), (([0910, 0920], lo_3), \mathsf{false})\})\},$$

we expect our variable assignments to be

$$\left.\begin{array}{rcl} LO^{lo_1}_{\mathsf{true}} & = & q[lo_1 = \mathsf{true} | \mathbf{OR}_{lo_1} = \mathsf{false}] \\ LO^{lo_2}_{\mathsf{false}} & = & q[lo_2 = \mathsf{false} | \mathbf{OR}_{lo_2} = \mathsf{true}] \\ LO^{lo_3}_{\mathsf{false}} & = & q[lo_3 = \mathsf{false} | \mathbf{OR}_{lo_3} = \mathsf{true}] \\ \mathbf{OR}^{lo_1}_{\mathsf{false}} & = & q[\mathbf{OR}_{lo_1} = \mathsf{false}] \\ \mathbf{OR}^{lo_2}_{\mathsf{true}} & = & q[\mathbf{OR}_{lo_2} = \mathsf{true} | lo_1 = \mathsf{true}] \\ \mathbf{OR}^{lo_3}_{\mathsf{true}} & = & q[\mathbf{OR}_{lo_1} = \mathsf{true} | lo_1 = \mathsf{true}, lo_2 = \mathsf{true}] \end{array}\right\} = 1$$

with all other variables being zero. Since the only variables which incrue costs are the $q[\ldots]$ variables, the cost of the assignment is $-\log(1/3) - \log(1) - \log(1) - \log(1) - \log(1) - \log(1) = -\log(1/3)$ and thus the probability of the assignment is $1/3$ as expected. As informally demonstrated in this example, the cost of a variable assignment is found by summing the product of each variable in ? and its corresponding cost in ψ.

Definition 14. *A* variable assignment *for a constraint system $L = (?, I, \psi)$ is a function s from ? to $\Re$. Furthermore,*

1. *If the range of s is $\{0, 1\}$, then s is a* 0-1 assignment*.*
2. *If s satisfies all of the constraints in I, then s is a* solution *for L.*
3. *If s is a solution for L and is also a 0-1 assignment, then s is a* 0-1 solution *for L.*

Definition 15. *Given a constraint system $L = (?, I, \psi)$, we construct a function Θ_L from variable assignments to $\Re$ as follows:*

$$\Theta_L(s) = \sum_{\gamma \in \Gamma} s(\gamma)\psi(\gamma, \mathsf{true}) + (1 - s(\gamma))\psi(\gamma, \mathsf{false}) \tag{18}$$

Θ_L *is called the* objective function *of* L.

Definition 16. *An* optimal 0-1 solution *for a constraint system* $L = (?, I, \psi)$ *is a 0-1 solution which minimizes* Θ_L.

By finding an optimal 0-1 solution for a constraint system, we find the most probable explanation for the corresponding PTN. Santos [29] presents a customized algorithm using the *cutting plane method* [26] for finding the optimal 0-1 solution. Since any Bayesian network can be represented as a PTN[5], we know that, in general, belief revision over PTNs is NP-hard [8, 27].

8 Summary and Conclusion

The probabilistic temporal network can represent very complicated and traditionally difficult domains. Our research has focused on exploring recurrence and periodicity, temporal spacing between cause and effect, and modeling the time-of-reference. These are traditional problems for temporal models. We are currently focusing our efforts on exploring these and other knowledge engineering issues.

In this paper, we introduced a constraint satisfaction formulation for performing belief revision. This formulation needs to be extended to perform belief updating (finding the most likely state of a given interval-RV pair or temporal aggregate). The constraint set needs to be enhanced to take better advantage of the structure imposed by our network structure. Future work will explore the possibilities of classes of networks with polynomial run-time behaviours. Other techniques such as path consistency may also be applicable to PTNs [37].

Aliferis and Cooper [1] have also developed a preliminary temporally extended Bayesian network formulation termed the *Modifiable Temporal Bayesian Network-Single Granularity* (MTBN-SG). A MTBN-SG is primarily an extended time-sliced Bayesian network defined over a range of time points. Each

[5]Treat each RV in the BN as a TA with a single interval-RV pair, using the ($\{=\}$, **PASSTHROUGH**) TCR, and make all intervals in the TAs equivalent.

ordinary node in a MTBN-SG is indexed over this entire range. Edges between nodes are represented by a *mechanism* variable which is a Boolean `true/false` random variable indicating whether the link is active, i.e. whether a dependency exists between the connected variables. Each such mechanism has an associated *lag* random variable defined over the range of time points indicating the delay between the "cause" and the "effect." Atemporal or *abstract* random variable nodes are supported which are not instantiated for each time point. The resultant graph can have cycles to allow expressions of recurrence and feedback. As long as all cycles in the underlying joint distribution have zero probability, the graph is said to be *well-defined*.

Since the edges, both *mechanism* and *lag* components, are represented by random variables, the edges can be both dependent on and causal to other random variables in the network. This allows the knowledge engineer to express conditions where a relationship exists between variables only under certain circumstances. The problem with this approach is that joint distributions can be described which are not compatible with the Bayesian model. Maintaining consistency in the local probability tables across random variables then becomes a concern.

In our work, we allow overlapping intervals so that events happening over intervals can be expressed. For example if a switch could be on from 1000 to 1030 or 1015 to 1045, this could be represented as $\{([1000, 1030], S_0), ([1015, 1045], S_1)\}$ where S_0 and S_1 are random variables for the switches position. S_1 would be conditioned on S_0 to prevent the switch from being on over both intervals. However, it is possible that the switch could be considered both on and off in the interval $[1015, 1030]$. To resolve this, one possibility is to make the interval itself random. For example $\{([1000, 1030], S_0), ([1015, 1045], S_1)\}$ might become $\{(I, On)\}$ where $P(I = [1000, 1030]) = P(S_O | \ldots)$ and $P(I = [1015, 1045]) = P(S_1 | \ldots)$. This gets us to only one interval.

In conclusion, we have presented a new knowledge representation for merging time and uncertainty. The technique, the probabilistic temporal network, draws from the independence semantics of Bayesian

networks and from the interval algebra. Methods for reasoning over the model have been introduced using techniques from operations research and from Bayesian belief revision.

References

[1] Constantin F. Aliferis and Gregory F. Cooper. A structurally and temporally extendend Bayesian belief network model: Definitions, properties, and modeling techniques. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI–96)*, pages 28–39, 1996.

[2] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[3] James F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154, 1984.

[4] James F. Allen and Johannes A. Koomen. Planning using a temporal world model. In *Proceedings of the Eight International Joing Conference on Artificial Intelligence*, pages 741–747, August 1983.

[5] Bruce G. Buchanan and Edward H. Shortliffe. *Rule-Based Expert Systems*. Addison Wesley, 1984.

[6] Eugene Charniak. Bayesian networks without tears. *AI Magazine*, 12(4):50–63, 1991.

[7] Eugene Charniak and Solomon E. Shimony. Probabilistic semantics for cost based abduction. In *Proceedings of the AAAI Conference*, pages 106–111, 1990.

[8] Gregory F. Cooper. Probabilistic inference using belief networks is NP-hard. Technical Report KSL-87-27, Medical Computer Science Group, Stanford University, 1987.

[9] Thomas Dean and Michael Wellman. *Planning and Control*. Morgan Kaufmann, 1991.

[10] Rina Dechter, Itay Meiri, and Judea Pearl. Temporal constraint networks. *Artificial Intelligence*, 49:61–95, 1991.

[11] Marek J. Druzdel. There is more than one way to skin a cat: Commentary on john lemmer's 'the causal markov condition: Fact or artifact?'. *ACM SIGART Bulletin*, 7(4):1–2, 1996.

[12] Christian Freska. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54:199–227, 1992.

[13] Avelino J. Gonzalez and Douglas D. Dankel. *The Engineering of Knowledge-Based Systems: Theory and Practice*. Prentice-Hall, Inc., 1993.

[14] Steve Hanks, David Madigan, and Jonathan Gavrin. Probabilistic temporal reasoning with endogenous change. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, August 1995.

[15] Jerry R. Hobbs, Mark Stickel, Paul Martin, and Douglas Edwards. Interpretation as abduction. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 95–103, 1988.

[16] John R. Josephson and Susan G. Josephson. *Abductive Inference: Computation, Philosophy, Technology.* Cambridge University Press, 1994.

[17] Keiji Kanazawa. *Reasoning about Time and Probability.* Thesis, Brown University, May 1992.

[18] Jak Kirman, Ann Nicholson, Moises Lejter, Eugene Santos, Jr., and Thomas Dean. Using goals to find plans with high expected utility. In *Proceedings of the Second European Workshop on Planning*, 1993.

[19] Uffe Kjaerulff. dhugin: A computational system for dynamic time-sliced bayesian networks. *International Journal of Forecasting, Special Issue on Probability Forecasting, to appear*, 1995.

[20] P. B. Ladkin and R. D. Maddux. On binary constraint problems. *Journal of the ACM*, 41(3):435–469, 1994.

[21] John F. Lemmer. The causal markov condition, fact or artifact? *ACM SIGART Bulletin*, 7(3):3–16, 1996.

[22] Gerard Ligozat. Weak representations of interval algebras. In *Proceedings of the AAAI Conference*, pages 715–720, 1990.

[23] Paul B. Losiewicz. Complexity, ontology, and the causal markov assumption. *ACM SIGART Bulletin*, 7(4):13–18, 1996.

[24] Zohar Manna and Amir Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: specification.* Springer-Verlag, New York, 1992.

[25] Drew V. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6:101–155, 1982.

[26] George L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors. *Optimization: Handbooks in Operations Research and Management Science Volume 1*, volume 1. North Holland, 1989.

[27] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA, 1988.

[28] Yun Peng and James A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving.* Springer-Verlag, 1990.

[29] Eugene Santos, Jr. On the generation of alternative explanations with implications for belief revision. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 337–347, 1991.

[30] Eugene Santos, Jr. On modeling time and uncertainty for diagnosis through linear constraint satisfaction. In *Proceedings of International Congress on Computer Systems and Applied Mathematics Workshop on Constraint Processing*, pages 93–106, 1993.

[31] Eugene Santos, Jr. Unifying time and uncertainty for diagnosis. *Journal of Experimental and Theoretical Artificial Intelligence*, 8:75–94, 1996.

[32] G. A. Shafer. *Mathematical Theory of Evidence.* Princeton University Press, 1979.

[33] Yoav Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, 1988.

[34] E. H. Shortliffe. *Computer-Based Medical Consultation: MYCIN.* Elsevier, 1976.

[35] E. H. Shortliffe and B. G. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[36] Patrick Suppes. *A probabilistic theory of causality.* Acta philosophica Fennica. Fasc. 24. North-Holland Publishing Co., Amsterdam, 1970.

[37] Peter van Beek and Dennis W. Manchak. The design and experimental analysis of algorithms for temporal reasoning. *Journal of Artificial Intelligence Research*, 4:1–18, 1996.

[38] Marc Vilain and Henry Kautz. Constraint propagation algorithms for temporal reasoning. In *Proceedings of the AAAI Conference*, pages 377–382, 1986.

[39] Joel D. Young and Eugene Santos, Jr. Introduction to temporal bayesian networks. In Michael Gasser, editor, *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference.* URL: http://www.cs.indiana.edu/event/maics96/Proceedings/Young/maics-96.html, April 1996.

[40] L. A. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, 11:199–227, 1983.